Functional evolution of the Colony Stimulating Factor 1 Receptor (CSF1R) and its ligands in 2 birds

DOI: 10.1002/JLB.6MA0519-172R

JLB JOURNAL OF LEUKOCYTE BIOLOGY

ARTICLE

Functional evolution of the colony-stimulating factor 1 receptor (CSF1R) and its ligands in birds

David A. Hume[1] | Maria W. Gutowska-Ding[2] | Carla Garcia-Morales[3] | Adebabay Kebede[4,5,6] | Oladeji Bamidele[7] | Adriana Vallejo Trujillo[8] | Almas A. Gheyas[2,9] | Jacqueline Smith[2,9]

[1]Mater Research Institute-University of Queensland, Translational Research Institute, Woolloongabba QLD 4102, Australia

[2]The Roslin Institute, University of Edinburgh Midlothian, United Kingdom

[3]Department Biotecnologia, Universidad Automona del Estado de Mexico, Toluca Area, Mexico

[4]Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa, Ethiopia

[5]Amhara Regional Agricultural Research Institute, Bahir Dar, Ethiopia

[6]International Livestock Research Institution (ILRI), Addis Ababa, Ethiopia

[7]African Chicken Genetic Gains Project-Nigeria, The International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

[8]Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

[9]Centre for Tropical Livestock Genetics and Health, University of Edinburgh, Midlothian, United Kingdom

Correspondence
David A. Hume, Mater Research Institute-University of Queensland, Translational Research Institute, Woolloongabba QLD 4102, Australia.
Email: David.Hume@uq.edu.au

Abstract

Macrophage colony-stimulating factor (CSF1 or M-CSF) and interleukin 34 (IL34) are secreted cytokines that control macrophage survival and differentiation. Both act through the CSF1 receptor (CSF1R), a type III transmembrane receptor tyrosine kinase. The functions of CSF1R and both ligands are conserved in birds. We have analyzed protein-coding sequence divergence among avian species. The intracellular tyrosine kinase domain of CSF1R was highly conserved in bird species as in mammals but the extracellular domain of avian CSF1R was more divergent in birds with multiple positively selected amino acids. Based upon crystal structures of the mammalian CSF1/IL34 receptor-ligand interfaces and structure-based alignments, we identified amino acids involved in avian receptor-ligand interactions. The contact amino acids in both CSF1 and CSF1R diverged among avian species. Ligand-binding domain swaps between chicken and zebra finch CSF1 confirmed the function of variants that confer species specificity on the interaction of CSF1 with CSF1R. Based upon genomic sequence analysis, we identified prevalent amino acid changes in the extracellular domain of CSF1R even within the chicken species that distinguished commercial broilers and layers and tropically adapted breeds. The rapid evolution in the extracellular domain of avian CSF1R suggests that at least in birds this ligand-receptor interaction is subjected to pathogen selection. We discuss this finding in the context of expression of CSF1R in antigen-sampling and antigen-presenting cells.

KEYWORDS

CSF1, ligand, macrophage, receptor, selection

1 | INTRODUCTION

Macrophage colony-stimulating factor (CSF1 or M-CSF) is a hematopoietic growth factor that regulates the survival, proliferation, and differentiation of mononuclear phagocytes.[1–3] CSF1 signals through a type III tyrosine kinase (TK) CSF1 receptor (CSF1R, also known as MCSFR, or c-Fms and recognized by anti-CD115 antibodies), which is expressed on the surface of macrophages, monocytes, and their progenitors. Since macrophages control many aspects of tissue regeneration, inflammation, and pathology, CSF1R signaling has been considered a target for the development of therapeutic agonists and antagonists (reviewed in Ref. 2). Loss-of-function (LOF) mutations in the *CSF1* and *CSF1R* loci in mice and rats are associated with depletion of blood monocytes and most tissue macrophage populations. The

Received: 31 May 2019 | Revised: 2 July 2019 | Accepted: 29 July 2019

phenotypic consequences differ depending on genetic background and species but include osteopetrosis and postnatal growth retardation.[4,5] Conversely, administration of CSF1 to mice, rats, or pigs produces a monocytosis and expansion of tissue macrophage populations.[6–8] In humans, gain-of-function coding mutations in *CSF1R* have been associated with an autosomal-dominant human neurodegenerative disease,[9,10] while two recent studies describe recessive loss-of-function *CSF1R* mutations[11,12] that share skeletal abnormalities with the mouse and rat *Csf1r* knockouts. Variants at the *CSF1* locus are strongly associated with Paget's disease.[13] Differences in phenotype of *Csf1r*$^{-/-}$ mice compared to a spontaneous *Csf1* mutation (*Csf1*$^{op/op}$) mice suggested the existence of a second CSF1R ligand, which was subsequently identified and named interleukin 34 (IL34).[14] Mutation of the *Il34* locus in mice revealed a specific function in development of subsets of tissue macrophages in skin and brain, where the gene is most highly expressed.[15] The two CSF1R ligands appear functionally equivalent. IL34 expressed under the control of the CSF1 promoter rescues the *Csf1*$^{op/op}$ phenotype.[16] The CSF1R system of two ligands binding to one receptor was shown to be conserved throughout vertebrates, including birds[17] and fish.[18] An intronic enhancer that controls CSF1R expression is also conserved from reptiles to mammals.[19] Recombinant CSF1 administered to chicks produced a massive expansion of blood and tissue macrophage populations.[20]

Solution of the tertiary structures of mouse and human CSF1 revealed the characteristic four alpha helices with two beta sheets, a structure shared by a large family of cytokines. The 3D structures of human/mouse IL34 also highlighted four antiparallel alpha helices, but with two shorter beta sheets partially replaced with an additional three alpha helices. Subsequent studies revealed the distinctive structures of the complexes between CSF1, IL34, and the receptor.[21–23] The CSF1R protein consists of five extracellular, Ig-like domains (D1-D5), a short transmembrane domain (TM), and an intracellular TK domain. The two N-terminal Ig-like domains (CSF1R$_{D2\text{-}D3}$) mediate ligand binding while the two extracellular Ig-like domains (CSF1R$_{D4\text{-}D5}$) are involved in receptor dimerization, which is required for downstream signaling (reviewed in Ref. 3).

Most immune proteins are subjected to an "arms-race" between host and pathogen and experience a strong positive selective pressure.[24,25] With some caveats,[26] nonsynonymous (amino acid altering) to synonymous substitution rate ratio (ω = dN/dS) provides a measure of natural selection at the protein level, where $\omega = 1$, $\omega > 1$, and $\omega < 1$ indicate neutral evolution, purifying, and positive selection, respectively.[27] The average dN/dS ratio of annotated immune-associated genes is up to four times higher than the genome-wide average for protein-coding genes.[24,25] Previous analysis on limited datasets indicated that both *CSF1* and *CSF1R* were subject to positive selection in birds, whereas *IL34* was subject mostly to purifying selection.[17]

Since the original characterization of the CSF1R system in chicken and zebra finch[17] the Avian Phylogenomic Consortium[28] completed the draft genome sequences for 48 bird species, representing all extant clades and many targeted projects since that time have further expanded the number of partial or complete genomes to >300 and the pool of predicted protein sequences for genes expressed in avian immune cells. Among many applications, these data permitted a re-evaluation of the gene content of avian genomes and global analysis of dN/dS ratios.[29] The expanded number of genomic sequences has added greatly to the diversity of avian predicted CSF1R, CSF1, and IL34 protein sequences. The current study takes advantage of the multispecies genomic dataset to examine the contrasting evolutionary constraints on the CSF1R system in birds and mammals.

2 | MATERIALS AND METHODS

2.1 | Sequence collection and multiple sequence alignment

Avian CSF1, IL34, and CSF1R protein and gene sequences were retrieved from the National Centre for Biotechnology Information (NCBI; http://ncbi.nlm.nih.gov) and completed avian genomes were analyzed by Avian Phylogenetic Consortium.[28] Accession numbers for all protein sequences are provided in Supplementary Table 4.

2.2 | Structural modeling

3D models of chCSF1, IL34, and CSF1R were created using YASARA program. Mouse CSF1 (pdb 3ejj-a) and human IL34 and CSF1R structures (pdb 4dkd-a and 4dkd-c) were used as templates for homology modeling. Next, avian models were compared with their mammalian equivalents using MUltiple STructural AligNment AlGorithm (MUSTANG) program. Ligand docking was performed using Autodock provided by YASARA.

2.3 | Phylogenetic analysis

An MSA for avian sequences was created using CLUSTALW and phylogenetic analysis generated by nearest neighbor joining (MacVector, Inc, www.macvector.com). Estimation of the ω rates of dN/dS amino acid substitution in avian CSF1, IL34, and CSF1R proteins was conducted using PAML v. 4.5 (Phylogenetic Analysis using Maximum Likelihood; http://abacus.gene.ucl.ac.uk/software/paml.html). For the CSF1R codon sequences, we ran the Fast Unconstrained Bayesian AppRoximation (FUBAR) module to estimate the dN/dS ratios for individual codons using default significance levels (HyPhy package via datamonkey.org). Further details of this analysis are provided in Ref. 30.

2.4 | Searching for protein coding variants in CSF1R, CSF1, and IL34 genes in commercial and village chicken populations

Sequence and variant data from a number of chicken populations—both commercial and noncommercial village chicken populations—were screened for the presence of protein coding variants from the genes CSF1R, CSF1, and IL34. Variant data on commercial broilers and layers originated from previous studies.[31,32] Sequence and variant data from a number of village chicken populations from diverse

climatic zones in Ethiopia (27 populations, 263 individuals) and Nigeria (14 populations, 122 individuals) were also screened. Sequence data on these African chicken populations have been generated as part of the African Chickens Genetic Gains project (https://africacgg.net/) and analyzed by researchers in the Centre for Tropical Livestock Genetics and Health (www.ctlgh.org). These African chicken samples were sequenced on the Illumina HiSeqX platform to produce paired-end reads with 30X mean coverage. Upon checking the sequence quality with the FastQC programme (v0.11.5), the sequence data were mapped against the GRCg6a reference genome using BWA-mem (v0.7.15) and variants called by applying the GATK (v3.8.0) Best Practice Guide for Germline Short Variant Discovery pipeline (https://software.broadinstitute.org/gatk/best-practices/workflow?id=11145). Full annotation of the whole genome sequences and analysis of genetic diversity of these chicken populations will be published elsewhere.

2.5 | Cross-species reactivity of recombinant chicken CSF1 proteins on growth and differentiation of avian BM-derived macrophage precursors

For each avian species examined (chicken, turkey, zebra finch, quail, and duck), 3-week-old birds were sacrificed and BM cells were obtained by flushing the marrow from two femurs and two tibias with PBS using a syringe and a blunt needle. For each preparation, cells were pelleted and resuspended in 4 mL complete RPMI (supplemented with 10% heat-inactivated FCS, 2 mM L-glutamine, 100 μg/mL penicillin, and 100 μg/mL streptomycin) containing 350 ng/mL recombinant chicken CSF1 (chCSF1). Cells were plated in 60 mm bacteriological plates (10^6 cells/mL) and incubated at 37°C in a CO_2 incubator for 8 days. After 4 days of incubation, fresh media supplemented with chCSF1 was added. Media containing the recombinant cytokine was exchanged every 2 days there afterwards. Biological activity was confirmed by visual examination of the plates for the formation of adherent macrophage cells and by harvesting and counting the cells. Each experiment was conducted in triplicate.

2.6 | Construction of chicken and zebra finch CSF1 receptor domain swapped CSF1 proteins

Four constructs were designed and synthesized by GeneArt (Life Technologies, Renfrew, UK). These synthetic DNA expression constructs encoded either the mature 189 amino acid long peptides of chCSF1 (M1-P189) and zfCSF1 peptides or the two domain swapped constructs, zfCSF1 containing the predicted chicken ligand binding Site 1 residues T86 to E111 and the chCSF1 containing the zebra finch residues K87 to N112 inclusive (note the construct includes the signal peptide and the numbering of amino acids refers to the full-length protein). All four expression constructs contained attB flanking regions for Gateway cloning (Life Technologies, Renfrew, UK). The "One-tube" protocol using the Gateway BP and LR Clonase II enzyme mix was used. Constructs were first introduced into an entry vector pDONR221 and then immediately into the destination expression vector pDEST51 containing V5 and 6xHis tags. Note that 50 μL of Top10 competent *Escherichia coli* was transformed with 1 μL of each construct using a standard transformation protocol. Bacterial colonies were expanded and DNA was extracted using EndoFree Plasmid Maxi Kit (Qiagen, Manchester, UK).

2.7 | Assay of the biological activity of chicken and zebra finch CSF1 proteins using growth factor dependent cells

We have previously established a bioassay for chicken CSF1 by stably transfecting the interleukin 3 (IL3)-dependent BaF3 cell line with a chCSF1R expression plasmid.[17] The transfected BaF3 cells express chCSF1R on the cell surface[33] and are able to survive and proliferate in the presence of chCSF1. HEK293 cells were transfected with expression plasmids for chCSF1, zfCSF1, zf_chCSF1, or ch_zfCSF1 and supernatants were collected. BaF3/chCSF1R cells were cultured with 20% HEK293T supernatant. A negative control had no added growth factor, while 5% conditioned medium from X63 Ag8-653 myeloma cells carrying an expression vector for IL-3 provided a positive control.[17] Cells were cultured at 2×10^4 cells per well in a total volume of 100 μL complete DMEM supplemented with an appropriate amount of HEK293 supernatant and grown for 48 h at 37°C in a 96-well plate. Viable cells were assayed as described previously.[17] Ten microliters of 3-(4 5-di methyl thiazol-2-yl)-2 5-diphenyltetrazolium) solution (final concentration 1 mg/mL) was added to each well and incubated at 37°C for 3 h. Then, 100 μL of solubilization solution (acid isopropanol) was added to each well and the absorbance was read at 570 nm.

2.8 | Production of recombinant avian CSF1 in HEK293 cells

HEK293T cells (American Type Culture Collection, Manassas, VA, USA) were cultured in DMEM (Sigma–Aldrich, Dorset, UK) supplemented with 10% HI-FCS (Sigma), 2 mM L-glutamine (Life Technologies, Renfrew, UK), 100 μg/mL streptomycin (Life Technologies, Renfrew, UK), 100 U/mL penicillin, and 0.1 mM nonessential amino acids (Life Technologies, Renfrew, UK). On the day before the transfection, 0.8×10^6 cells/well in a six-well plate were plated with antibiotic-free DMEM followed by transfection with 4 μg DNA (chCSF1_pDEST, zfCSF1_pDEST, ch_zfCSF1_pDEST, zf_chCSF1_pDEST, or empty pDEST51 DNA) using Lipofectamine 2000 (Life Technologies, Renfrew, UK). Supernatant containing secreted protein was harvested after 48-h incubation at 37°C with 5% CO_2. Protein expression was confirmed using antibodies against the C-terminal His tags to visualize expressed peptides on a Western blot.[17]

3 | RESULTS

3.1 | Sequence analysis of the CSF1 ligand-receptor system from birds and mammals

From available genomic DNA sequences and entries in NCBI GenBank, we were able to extract 68 CSF1R, 30 IL34, and 36 CSF1 predicted full-length protein sequences orthologous to the functional chicken

proteins analyzed previously.[17] The relative paucity of avian CSF1 and IL34 sequences available reflects the difficulties in sequencing in the respective genomic regions, in common with multiple other GC-rich regions, in all avian genomes.[29] In many cases, the sequences annotated as CSF1 or IL34 in NCBI as a predicted protein were truncated at the N terminus relative to full-length chicken and zebra finch orthologs. Multiple sequence alignments (MSAs) of each of the avian CSF1, IL34, and CSF1R protein-coding regions are provided in Supplementary Table 1A–C. In mammals, the *CSF1* locus encodes multiple isoforms of the protein generated by alternative splicing.[3] The longest cDNA encodes a membrane-bound precursor that is cleaved from the cell surface by TNF-alpha converting enzyme (TACE, ADAM17)[34] to release the minimal bioactive CSF1 protein. In transgenic mice, this longer form of the ligand is required to fully complement a CSF1 mutation and restore postnatal growth.[35] Consistent with previous evidence of the production of longer forms of *CSF1* mRNA in chickens,[17] some avian genomes have been shown to encode CSF1 proteins of 450–550 amino acids but such longer sequences constitute only a small subset of the currently available CSF1 protein sequences on NCBI GenBank. Alignment of the chicken (XP_0154359; isoform X1) and zebra finch (ACS32142) longest CSF1 amino acid sequences with mammalian (mouse and human) proteins (Fig. 1) reveals limited conservation of the bioactive peptide between birds and mammals. Birds and mammals also encode a TM. The short intracellular domain contains a membrane-proximal basic region that is conserved between mammals and birds. The remainder of the intracellular domain is also strongly conserved in birds. Similar membrane proximal basic domains are found in many membrane-associated proteins including G protein-coupled receptors. The intracellular domain may function to promote membrane trafficking from the Golgi[36] or conceivably also produce a reverse signal to the CSF1-producing cell.[37] The intervening region between the bioactive peptide and membrane is longer in mammals than in birds. In common with many proteolytic cleavage domains, the obvious conserved feature is repeated proline (P), glutamate (E), serine (S), and threonine (T) amino acids.

At the N terminus, we also noted that there was considerable ambiguity among predicted protein sequences in GenBank regarding the location of the start codon and the length of the leader sequence. For the purpose of the current analysis, we have aligned the processed peptide containing the 160 amino acids that make up the minimal bioactive 4-helix bundle.[17] In the case of IL34, the predicted avian proteins are all around 180 amino acids, truncated at the C-terminus relative to predicted mammalian IL34 proteins (230–240 amino acids). In mammals, some of the C-terminal amino acids were found to be engaged in binding to CSF1R[23] but in birds the 180 amino protein contains the biological activity.[17]

As noted based upon comparison of chicken and zebra finch,[17] the avian CSF1 sequences all showed conservation of cysteines that provides a strong reference framework for the alignment (Supplementary Table 1A). These conserved avian residues are predicted to form three intrachain disulfide bonds coincident with the cysteines involved in disulfide bonds in CSF1 of mammals and fish.[17] In all of the avian CSF1 peptides, the cysteine responsible for the interchain disulfide bond in mammalian CSF1 is substituted with glycine (G29 in Supplementary Table 1A; position 63 in Fig. 1). Nevertheless, the chicken protein forms a dimer through predicted large hydrophobic interfaces.[17] Early studies indicated that the interchain disulfide in human CSF1 was absolutely required for dimerization and biological activity, but this does not appear to be the case.[38] Mutation of this cysteine (C31S, numbered in the mature CSF1 peptide without the leader sequence) did not compromise refolding or biological activity of recombinant human CSF1. Based upon structural analysis, two amino acids (Q26 and M27) were predicted to make strong contributions to dimer formation.[38] These are conserved in all bird and mammalian CSF1 sequences (Q25/M26 in the active mature chicken sequence shown in Supplementary Table 1A; positions 58/59 in Fig. 1). Indeed, D23, which made strong electrostatic and nonpolar contributions to the dimer interface in the C31S mutant human protein, is also conserved between birds and mammals and in all birds (Supplementary Table 1C). A second shorter segment in CSF1 that contributed to the dimer interface, R66-N73 in human CSF1 (positions 98–107 in Fig. 1), is also conserved between mammals and birds and the core (FKENS) is identical in all bird species. A combined C31S/M27R mutation produced a monomeric CSF1 that acted as a CSF1R antagonist. The absence of cysteine in this location in the avian ligand suggests that the C31S mutation in the mammalian protein is unlikely to be necessary to achieve this outcome. Our earlier analysis of available CSF1 sequences indicated significant divergence among species and evidence of positive selection.[17] This conclusion was confirmed using the larger dataset.[30] Figure 2 shows a neighbor-joining phylogenetic tree for the available sequences. This simple analysis reveals that the Galloanseriformes (chicken, turkey, guinea fowl, quail, and goose) clearly form a separate group.

Avian IL34, unlike CSF1, is subject to purifying selection.[17] Indeed, although CSF1 is highly divergent between birds and mammals, the core 145 amino acid chicken IL34 protein, excluding the leader sequence, is around 60% identical to the human protein and can be readily aligned (not shown). Despite this level of conservation, amino acid differences among mammalian species were associated with species-specific biological activity.[39] Supplementary Table 1B shows the MSA of the available avian IL34 proteins.

The intracellular domain of CSF1R including all of the tyrosines that undergo phosphorylation to initiate signaling is conserved from birds to mammals. MSA of avian CSF1R peptides reveals strong conservation of the transmembrane and intracellular TK domains (Supplementary Table 1C). The catalytic amino acids required for TK activity, the tyrosine residues that are autophosphorylated in response to receptor ligation,[3] and those involved in autoinhibition revealed by the crystal structure of the kinase domain of mammalian CSF1R[40] are conserved in all birds. This conservation is consistent with functional studies in which the chicken CSF1R was able to signal when expressed in mammalian growth factor-dependent cells.[17] Our previous study noted that a critical amino acid in the kinase domain, C665 in human, is substituted with arginine in birds. This amino acid contributes to the binding of the widely used CSF1R kinase inhibitor, GW2580 (Patent US20040002145; 2004), which was shown to be

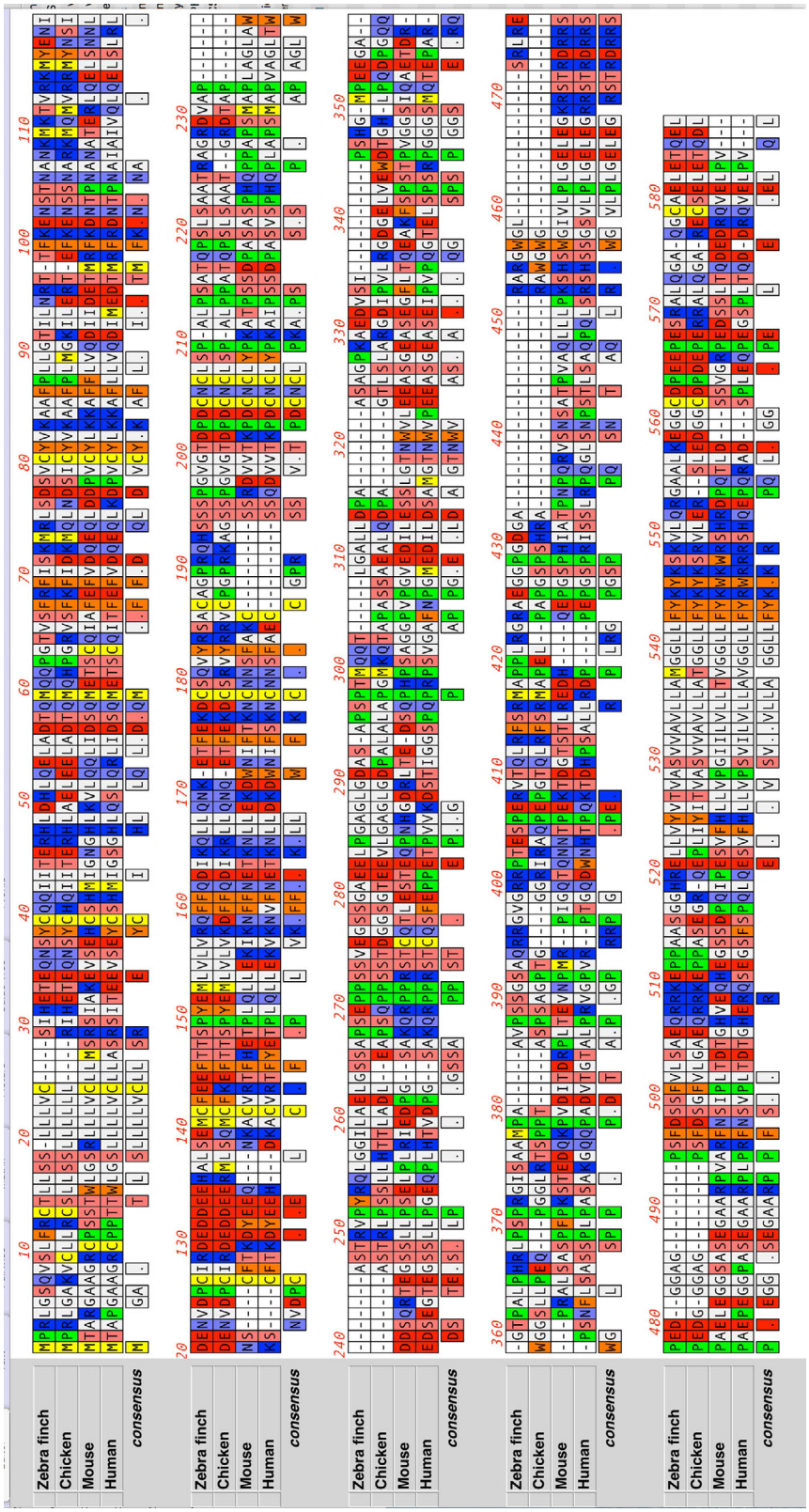

FIGURE 1 Alignment of long forms of CSF1 from birds and mammals. Protein sequences of 480 amino acid CSF1 sequences from chicken and zebra finch and 552 amino acid sequences from mouse and human were retrieved from NCBI. Multiple sequence alignment was performed using MUSCLE, in the MacVector programme. Features discussed in the text include the signal peptide (1–35), bioactive CSF1 peptide (35–196), conserved cysteines (39, 80, 126, 141, 178, 186), nonconserved mammalian-specific cysteine (63), PEST domain (200–520), transmembrane domain (520–542), and conserved membrane proximal basic domain (544–552).

inactive on the chicken receptor.[17] The arginine substitution is present in all bird sequences.

The topology of CSF1R extracellular domains is conserved and the cysteine residues in the five predicted Ig-like domains (D1–D5) in the extracellular domain of CSF1R are present in all available bird sequences. This provides a clear framework for structure-based alignment that is almost gap free (Supplementary Table 1C) and construction of a phylogenetic tree based upon the aligned sequences as shown in Fig. 3. The tree closely resembles the recently published comprehensive phylogeny based upon targeted next-generation sequencing of

FIGURE 2 **Phylogenetic analysis of CSF1 divergence among avian species.** Multiple sequence alignment of the bioactive 160 amino acid CSF1 molecule was performed using ClustalW in the MacVector package to generate the alignment in Supplementary Table 1A. A neighbor joining phylogenetic tree was then generated using the same package.

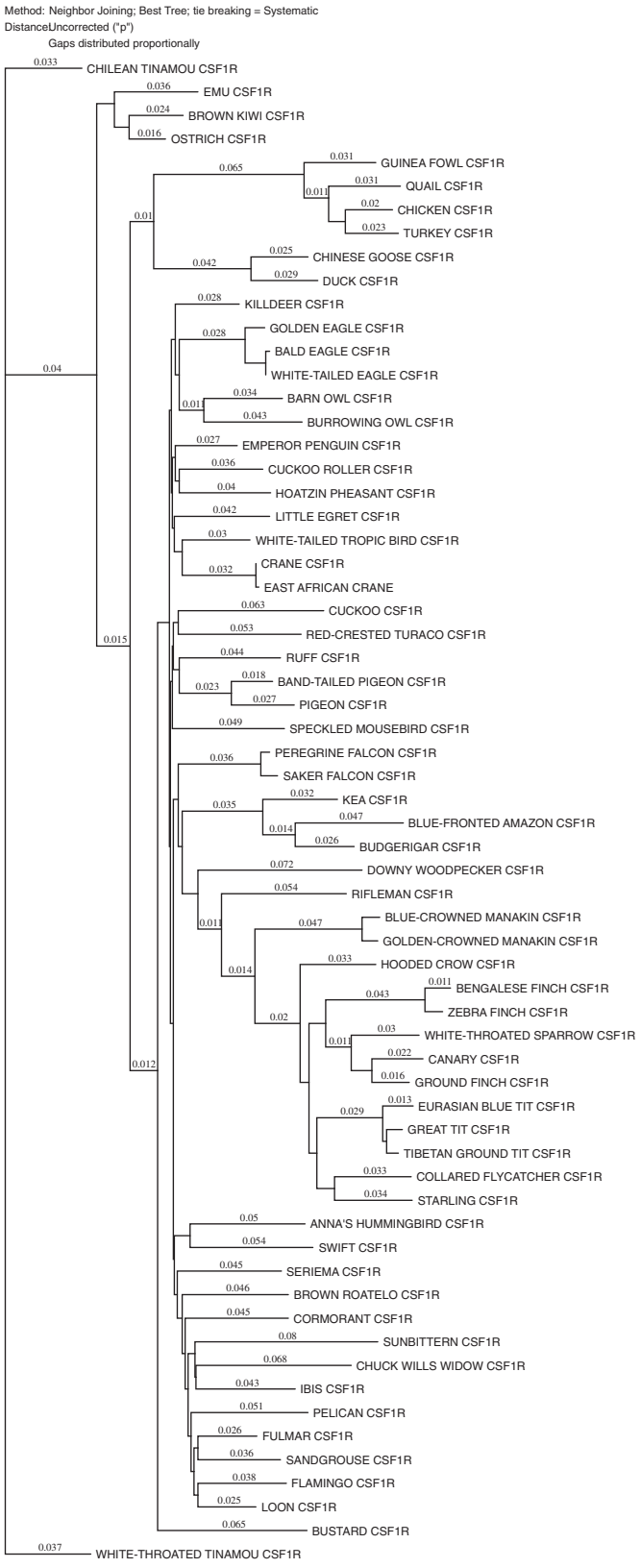

FIGURE 3 Phylogenetic tree analysis of CSF1R divergence among avian species. Multiple sequence alignment of the 968 amino acid CSF1 molecule was performed using ClustalW in the MacVector package to generate the alignment in Supplementary Table 1C. A neighbor joining phylogenetic tree was then generated using the same package.

a much larger assembly of avian species[41] (see the phylogenetic tree image from this study in the graphical abstract, reproduced with permission) and recapitulates analysis based upon the divergence of the conserved intronic enhancer in the CSF1R locus.[19] As in the case of CSF1, the Galloanserae form a divergent group.

The overall sequence identity between the most disparate CSF1R protein sequences (e.g., between chicken and zebra finch), around 75%, is similar to the conservation between the most divergent mammalian sequences (primates and rodents[39]) but the pattern of variation among species is different. The majority of substitutions/insertions among bird species occur in the region between domains 3 and 4, and in a small number of hypervariable regions. We repeated the dN/dS analysis using the larger dataset now available. Figure 4 shows a comparison of the profiles for avian and mammalian CSF1R relative to the predicted domain structures. A total of 15 amino acids showed $\omega > 1$ in the avian lineages.

3.2 | Cross-species specificity of the CSF1 ligand in birds

The tertiary structures of CSF1, IL34, and CSF1R in the chicken were modeled previously based upon published mammalian structures. The results were consistent with strong conservation of the tertiary structure.[17] Subsequently crystal structures of mammalian CSF1R/CSF1 and CSF1R-IL34 complexes have been reported.[21–23,42,43] In each case, ligand binding involves interaction with two sites (Sites 1 and 2) in the Ig-like domains D2 and D3 of the receptor. We recreated the 3D models of chCSF1, IL34, and CSF1R and the ligand-receptor complexes using Yet Another Scientific Artificial Reality Application (YASARA) program. Mouse CSF1/CSF1R (pdb 3ejj-a) and human IL34/CSF1R structures (pdb 4dkd-a and 4dkd-c) were used as templates for homology modeling. The avian models were compared with their mammalian equivalents using MUltiple STructural AligNment AlGorithm (MUSTANG) program. The modeling results are presented in detail in Ref. 30. In overview, these analyses confirmed that the ligand-receptor interfaces for the two ligands are likely to be positionally conserved from birds to mammals.

Based upon the structural analyses, we predicted the candidate Sites 1 and 2 regions of interaction for the avian CSF1/CSF1R and IL34/CSF1R complexes. The alignments of these regions from the most divergent chicken and zebra finch sequences and the corresponding human sequences are provided in Table 1 and the predicted amino acid interactions are summarized in Table 2. The binding modes clearly differ between birds and mammals. None of the positively charged amino acids in mammalian CSF1R (human R142, R144, R146, R150, K168) that contribute to shared interactions with CSF1 and IL34 in Site 1 on Domain 2[23] is conserved in birds and the corresponding amino acids in the predicted CSF1R structure are not conserved between zebra finch and chicken. Strikingly, several positively selected amino acids in Fig. 4 that fall within CSF1R domains 2 and 3 (positions 145, 147, 148, and 248 in chicken) also lie within hypervariable regions of CSF1R sequence divergence among avian species and impact on the contact amino acids within both Sites 1 and 2 (Supplementary Table 1C). Site 1

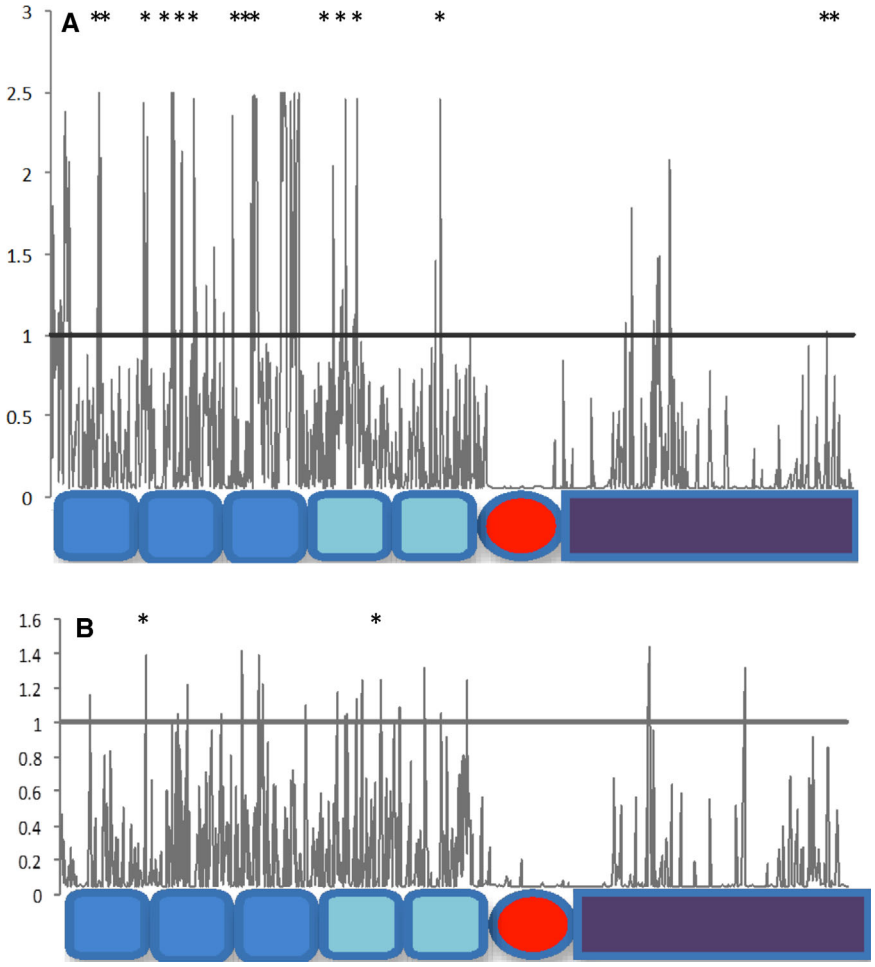

FIGURE 4 CSF1R evolutionary selection profile. ωvalues (Y) for individual amino acids (X) in (**A**) avian and (**B**) mammalian CSF1R, obtained in codeml and FUBAR analyses (performed as described in Materials and Methods section). A line across the diagram indicates $\omega = 1$ (neutral selection). Asterisks above the graphs indicate residues identified in FUBAR analysis as being positively selected with a P value < 0.05. The tertiary structure of the receptor is indicated by the dark blue squares (Ig domains D1–D3), light blue squares (Ig domains 4 and 5), red circle (transmembrane domain), and purple (intracellular tyrosine kinase domain).

divergence also distinguishes chicken, quail, and turkey from duck and goose. The structure-based alignment of predicted contact residues in CSF1 reveals corresponding variation in Site 1 of the ligand, in particular multiple nonconservative substitutions between chicken T57 and E82, whereas Site 2 on CSF1 is conserved across all available avian sequences. The Site 1 interaction between chCSF1 and chCSF1R is predicted to involve a salt bridge between K73 in the ligand and E168 and E170 in the receptor (Table 2). This interaction is abolished in the zebra finch receptor (Q164, S166); substitutions shared by many bird species (Supplementary Table 1C). Conversely, the zebra finch ligand has two charged amino acids, E60 and E63, that are likely to form salt bridges with R142 and K147 in chicken CSF1R (similar salt bridges exist between CSF1 E62/D63 and CSF1R R142/R146 in the human CSF1-CSF1R complex).

Felix et al.[43] also identified the structural basis for ligand-induced dimerization involving amino acids in Ig-like domain D4 of the receptor. In the human receptor, this interaction involves salt bridges between R370 and E375 across the dimer interface, a mechanism that is conserved among type III TK receptors. This region (highlighted in Supplementary Table 1C) can be aligned with mammalian CSF1R and the corresponding glutamate (E) residue is conserved across most bird species. The precise location of basic amino acids that contribute to dimerization may vary among the species.

In mammals, sequence variation in the ligand binding sites of the CSF1/CSF1R complex constrains cross-species reactivity.[39] For example, mouse CSF1 is not active on the human CSF1R, but is active on pig, while human and pig CSF1 are active on all three species.[39] Based upon the sequence analysis and structural models, we predicted that there would also be major constraints on cross-species reactivity of CSF1 between the most divergent species exemplified by chicken and zebra finch. To test this hypothesis and indirectly confirm the accuracy of the structural model, bone marrow (BM) cells from several avian species were treated with recombinant chicken CSF1 (chCSF1). This system has been used before to generate bone marrow-derived macrophages (BMDM) from the chicken for functional studies. Chicken BMDM generated in response to CSF1 were actively phagocytic, responded to lipopolysaccharide, and expressed macrophage-specific transcripts detectable by RNAseq.[17] Despite the variation in D2 and D3 between these species, chCSF1-stimulated cell survival and proliferation of BM cells from chicken, quail, turkey, and also duck and gave rise to a monolayer of macrophage-like cells within 7 days. By contrast, zebra finch BM cells died when cultured with chCSF1 with no mature macrophage-like cells present in the wells after 7 days (not shown). To test whether the amino acid differences in Site 1 of CSF1 were sufficient to explain the species specificity, we generated expression constructs encoding chCSF1, zfCSF1, and two domain

TABLE 1 Sequence divergence between chicken and zebra finch in the interaction sites between CSF1R and its ligands, CSF1 and IL34

CSF1																			
Site I																			
Human	F55*	Q58	D59	I60	M61*	E62*	D63*	M65	R66	A74	I75*	A76	V78*	Q81	E82	L85*			
Chicken	P53	G56	**T57**	I58	L59	**N60**	R61	**T63**	F64	**N72**	K73	M74	**K75**	R78	K79	E82			
Zebra finch	P53	G56	**K57**	I58	L59	**E60**	R61	**E63**	F64	**R72**	K73	M74	**Q75**	R78	R79	N82			
Site II																			
Human	Y6*	H9	M10	G12	S13*	G14	H15	R79*											
Chicken	Y4	Q7	I8	T10	E11	R12	H13	R78											
Zebra finch	Y4	Q7	I8	T10	E11	R12	H13	R78											
CSF1R-CSF1																			
Site I																			
Human	Q113	E114	E119	R142	V143*	R144	G145*	R146	P147*	L148*	M149*	R150	H151*	K168	F169*	I170*	S172		
Chicken	F113	R114	K119	**R142**	**N143**	**D144**	**G145**	**S146**	**K147**	L148	**S149**	P150	G151	E168	H169	E170	K172		
Zebra finch	F109	R110	K115	**M138**	**E139**	**N140**	**A141**	**P142**	**S143**	L144	**P145**	P146	G147	Q164	N165	S166	K168		
Site II																			
Human	V229	D230	V231	N232	F233	D234	V235	Q248*	Q249*	S250	D251	F252	H253	N254	N255	Y257	K259	S281	
Chicken	P229	S230	H231	K232	Y233	D234	I235	**G248**	**K249**	**P250**	**D251**	I252	**Y253**	D254	G255	Y257	I259	N281	
Zebra finch	P225	S226	H227	K228	Y229	D230	I231	**K244**	**M245**	**A246**	**G247**	L248	**E249**	N250	D251	Y253	I255	N277	
IL34																			
Site I																			
Human	S100	E103	S104	Q106	D107	L109	E111	W116	E123	T124	L125	L127	Q131	T134					
Chicken	Q90	L93	**H94**	**L96**	E97	L99	E101	**R106**	**S113**	Q114	L115	D117	**V121**	S124					
Zebra finch	Q90	L93	**R94**	**R96**	E97	L99	E101	**W106**	**C113**	Q114	L115	D117	**E121**	S124					
Site II																			
Human	T36	F40*	D43	K44	V71	E111	F72												
Chicken	E26	L30	D33	K34	V61	E91	L62												
Zebra finch	E26	L30	D33	K34	V61	E91	L62												
CSF1R-IL34																			
Site I																			
Human	E119	R142	V143*	R144	G145*	R146	P147*	L148	M149*	R150	H151*	K168	F169	I170*	S172	Q173*	D174	Q176	
Chicken	K119	**R142**	**N143**	**D144**	**G145**	**S146**	**K147**	L148	**S149**	P150	G151	**E168**	**H169**	E170	K172	G173	Q174	Q176	
Zebra finch	K115	**M138**	**E139**	**N140**	**A141**	**P142**	**S143**	L144	**P145**	P146	G147	**Q164**	**N165**	S166	K168	G169	S170	R172	
Site II																			
Human	V229	D230	V231	N232	F233	D234	A245	I246	P247	Q248	Q249	S250	D251	F252	H253	N254	N255	Y257	
Chicken	P229	S230	H231	K232	Y233	D234	K245	**S246**	N247	**G248**	**K249**	P250	**D251**	I252	Y253	D254	G255	Y257	
Zebra finch	P225	S226	H227	K228	Y229	D230	K241	**R242**	T243	**K244**	**M245**	A246	**G247**	L248	E249	N250	D251	Y253	

For the predicted CSF1-CSF1R interaction, the binding Sites 1 and 2 are based upon structure-based alignment of available human and mouse CSF1-CSF1R (D1-D3) and IL34-CSF1R (D1-D3) structures. Contact amino acids in CSF1 and CSF1R derived from the human structures are highlighted in gray, and asterisks indicate amino acids that differ between human and mouse. Where the corresponding amino acids diverge between zebra finch and chicken, they are set in bold.

swaps in which the variable regions of Site 1 in zebra finch CSF1 (K57-N82) was replaced with the homologous residues in chicken CSF1 (T57-E82) (zf_chCSF1) and vice versa (in zebra finch CSF1 (ch_zfCSF1). Four constructs were expressed in HEK293T cells and supernatants containing recombinant CSF1 were tested. The supernatants from HEK293 cells transfected with zfCSF1 expression plasmid were able to promote survival of BaF3 cells expressing the chicken CSF1R to the same extent as supernatants from cells expressing chCSF1 (Fig. 5). Both of the domain-swapped constructs zf_chCSF1 and ch_zfCSF1 were also active on the chCSF1R reporter cell line (Fig. 5A). We then tested the ability of each of the supernatants to promote the differentiation of chicken BM and zebra finch BM. The zebra finch BM cells survived and clearly differentiated to form a monolayer of macrophage-like cells in response to zfCSF1, whereas in chCSF1 no live cells remained in the cultures after 7 days. The domain swap in which chCSF1 contains K57-N82 from zfCSF1 (ch_zfCSF1) was also active on zebra finch marrow (Fig. 5B). By contrast, the reciprocal domain swap in which the zebra finch ligand contains T57-E82 from

TABLE 2 Amino acid interactions in the binding of chicken and zebra finch CSF1 to CSF1R

	CSF1		CSF1R	
	Chicken	Zebra finch	Chicken	Zebra finch
Site 1	L54	L54	**S146**	**P142**
	G56	G56	**S146**	**P142**
	L59	L59	**S146**	**P142**
	N60	**E60**	**R142, S146, K147,** P150	**M138, P142, G143,** P146
	T63	**E63**	P150	P146
	K73	**K73**	G151, **E168, E170**	G145, **Q164, S166**
	M74	**M74**	**L148,** G151, **E168**	**P144,** G147, **Q164**
	R78	R78	**S149**	**P145**
	K79	**R79**	**H169**	**N165**
	E82	**N82**	**E170**	**S166**
Site 2	Q7	Q7	H231, **P250**	H227, **A246**
	I8	I8	H231, Y257	H227, Y253
	E11	E11	**D251**	**G247**
	R12	R12	**D251, I252**	**G247, L248**
	H13	H13	**I252**, Y257	**L248**, Y253

The chicken and zebra finch CSF1R complexes were modeled based upon the human CSF1-CSF1R (D1-D3) structure as described in Materials and Methods section. Non-conserved amino acids are set in bold.

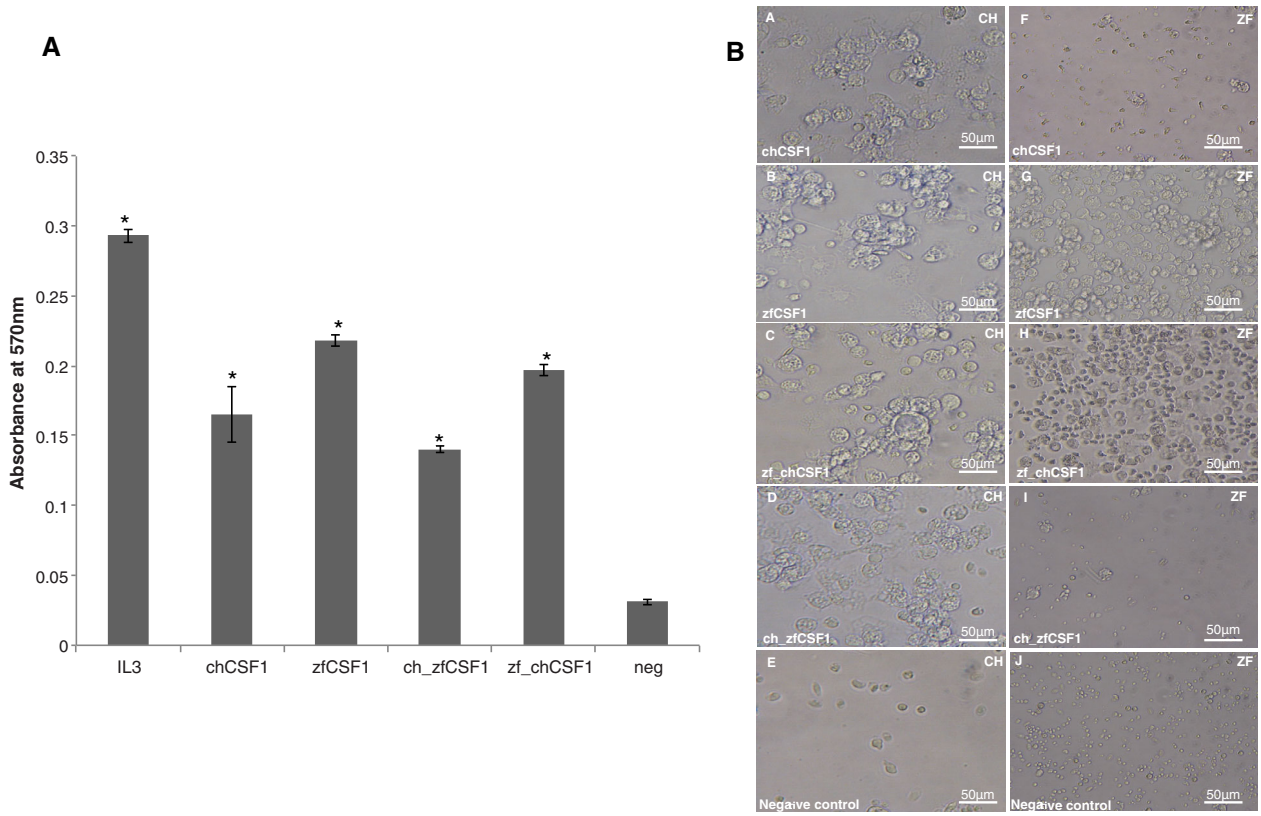

FIGURE 5 **Analysis of cross-species reactivity of chicken and zebra finch (ZF) CSF1.** (**A**) CSF1-dependent BaF3/chCSF1R cells were incubated for 48 h with supernatants from transfected HEK293 cells expressing recombinant chCSF1, zfCSF1, or domain swap proteins zf_chCSF1 and ch_zfCSF1 as described in Materials and Methods section. Interleukin 3 was included as a positive control and culture medium from untransfected HEK293 cells as a negative control. The absorbance (at 570 nm) is a measure of viable cell cells. Values are the average of three separate experiments and are expressed as mean ± SD. $^*P < 0.05$. (**B**) Fresh chicken (CH; at left) or ZF bone marrow cells were incubated with the same set of HEK293 supernatants for 1 week. In both cases, no cells survived in the absence of added growth factor (panels E and J). As shown in images in panels A–D, chicken bone marrow cells produced a relative confluent lawn of macrophages in response to all of the supernatants. Conversely, only zfCSF1 or zf_chCSF1 (chCSF1 with ZF Site 1) directed macrophage proliferation and differentiation from ZF marrow (panels G and H). Images are representative of three separate experiments.

Site 1 of the chicken ligand (ch_zfCSF1) abolished the activity on zebra finch marrow. This observation confirms that the difference in cross-species reactivity between chicken and zebra finch CSF1 ligands can be attributed to the variation in receptor binding Site 1 (Table 1).

3.3 | Polymorphism in the CSF1R, CSF1, and IL34 genes among selected chicken populations

Western commercial chickens have been subject to intensive selection of production traits: rapid growth and meat production or egg laying. Selection has produced genomic signatures that can be detected as extended regions of homozygosity.[44] In mammals, mutations in CSF1 or CSF1R produce severe postnatal growth retardation suggesting a link between macrophages and the growth hormone/IGF1 axis.[3,5] Indeed, the CSF1R gene on chromosome 13 lies within an interval containing signatures of selection[44] and 42 separate quantitative trait loci (QTL) (https://www.animalgenome.org/QTLdb) almost all associated with growth-related traits. We examined the impact of sequence variants among commercial birds (Supplementary Table 2). Protein-coding variants affecting CSF1, CSF1R, and IL34 were detected in pedigree lines of commercial broilers and layers sequenced to produce the avian high-density SNP chip.[31,32] Variants in CSF1 and IL34 were detected with minor allele frequencies (MAF) > 0.2 in specific lines of layers and broilers. None of the variants in CSF1 affects contact amino acids but V45G, L54S, V121G, F124H, and L132M each alter amino acids that are conserved across all other bird species (Supplementary Table 1A). Although N87D is recorded as a variant in one broiler line, all other avian sequences have an asparagine (N) in this position and this is clearly also the common sequence in commercial birds. N87D is predicted to be a deleterious substitution but our original chicken CSF1 cDNA[17] encodes D87 and was expressed as an active protein. In the case of CSF1R, consistent with evidence of selective sweeps in the genomic region, several protein-coding variants were detected with high MAF in specific broiler or layer lines. One variant, G414S, was universal in commercial birds (and also in African birds, below). Duck and goose each have a glycine (G) in this position, whereas turkey, quail, and guinea fowl have serine (S). So, this variant is probably a difference distinguishing domestic chickens from red jungle fowl.

A different selection pressure including heat stress and disease applies to indigenous chicken ecotypes selected for resilience and survival in tropical small holder systems.[45] We predicted that genes such as CSF1 and CSF1R that diverge rapidly between species might also exhibit functional polymorphism within a species occupying many diverse environmental niches. We therefore explored genomic DNA sequences from tropically adapted village chicken populations from multiple climatically diverse regions of Ethiopia and Nigeria. The results are shown in Supplementary Tables 3. Within the CSF1R gene in Ethiopian birds, we identified 13 nonsynonymous protein-coding variants with prevalent allele frequencies, all but one within the extracellular domain. Five of these variants were unique to the Ethiopian birds, whereas others were previously assigned SNP IDs during development of the 600,000 high-density SNP chip and have useful heterozygosity across commercial broiler and layer lines.[31,32] F125L (common to broilers and layers); N153S, S409L, and I468L (broiler-specific); and R294W and A308T (layer-specific) were also detected with variable frequency in most Ethiopian and Nigerian populations, perhaps reflecting admixture of Western birds. Among the variants shared by commercial and tropically adapted birds, none impacted amino acids implicated in ligand binding identified in Table 1 and most vary to some extent between species. Position F125 is also leucine (L) in most other avian species; position N153 is serine (S) in two species of tit, starling, and ruff and position 308 is threonine (T) in 2 manakin species (blue-crowned and golden-crowned) and glycine (G) in cuckoo roller (Supplementary Table 1C). Positions 294, 409, and 468 have more than one substitution across species, including the chicken variant.

One of the novel coding variants in tropically adapted birds, D91N, was detected in 26/27 Ethiopian populations with an average allele frequency of 0.32 (range 0.05–0.67). The same variant was detected in around half of the Nigerian populations with a lower MAF (average 0.08). This amino acid lies within a region of D2 that is conserved across bird species (Supplementary Table 1C). The D to N substitution requires a single base change but that substitution is predicted in only one other species, the Kea. Accordingly, we suggest that this change is due to some form of positive selection among the tropically adapted chickens in Ethiopia. Only five coding variants were detected in IL34 from African populations, of which only one, R127Q, was prevalent in multiple Ethiopian populations (Supplementary Table 3). This amino acid is conserved in bird species but lies outside the binding site for the receptor. In the biologically active portion of CSF1, we identified the N87D variant discussed above at low allele frequency in the majority of populations and a small number of rare potentially deleterious variants at low frequency in specific populations (Supplementary Table 4). None of the variants altered contact amino acids. One other variant detected in all Ethiopian populations, E99K, is also present in duck and goose, but not in quail or guinea fowl reference sequences.

4 | DISCUSSION

We have combined structural predictions and evolutionary analysis based upon a large collection of avian genomic sequences to identify selection pressure on CSF1R and its ligands in avian and mammalian lineages. This analysis addresses a quite fundamental question. How can the specificity of interaction between a single receptor and two ligands be maintained through evolution in the face of pathogen selection? We might have anticipated conservation of key amino acids involved in ligand recognition, but where the topology of interaction is conserved, the precise contacts have changed. Of course, the contact amino acids would also be targeted by pathogen-associated molecules that interfere with receptor activation. Our analysis revealed positive selection in avian lineages of multiple amino acids that are predicted to contribute to the recognition of CSF1 by CSF1R and corresponding changes in the ligand. Surprisingly, despite the extensive changes and contrary to our previous structure-based prediction,[17] zfCSF1 retained the ability to activate chCSF1R, whereas chCSF1 was inactive

on zfCSF1R. Domain swap analysis confirmed that the amino acids K57-N82 within zfCSF1 (Site 1) that interact with domain D2 of CSF1R are both necessary and sufficient to enable activation of zebra finch BM cells. There are six amino acid differences between the two species in this short segment, all involving charged amino acids (Table 1). As discussed above, we suggest that the binding affinity of chicken CSF1 for chicken CSF1R depends upon charged amino acid interactions. By contrast, there appear to be no predicted salt-bridge interactions in zebra finch CSF1 binding to its receptor, but two charged amino acid substitutions may permit the formation of salt bridges to the chicken receptor.

The analysis of many more *IL34* sequences in birds (Supplementary Table 1B) confirmed that unlike CSF1, this protein is under strong negative selection. Notwithstanding the high conservation in mammals, human IL34 is not active on the mouse CSF1R, and vice versa, but both were equally active on pig CSF1R.[39] Aside from the conserved basic amino acid contacts that are shared with CSF1, many of the orthologous positions to the contact amino acids bound by IL34 in human CSF1R also vary between avian species.[23] Although IL34 is highly conserved overall across avian species, the variation that does exist is focused in several of the contact amino acids in Site 1 of IL34, and there are corresponding changes in predicted contact amino acids in domain 2 (Site 1) of CSF1R (Supplementary Table 2). The strong negative selection acting on other parts of the IL34 sequence thus appears unrelated to binding to CSF1R. Indeed, there is evidence for IL34 interaction with two separate partners in mice, a protein tyrosine phosphatase[46] and syndecan-1.[47] The function of IL34 in birds has not been studied beyond the demonstration that the protein is active on the chicken CSF1R.[17]

The most striking feature of our analysis, which clearly distinguishes birds from mammals, is the hypervariability of the CSF1/IL34 binding Site 1 in CSF1R. Why has selection in avian evolution apparently acted upon ligand binding to CSF1R? One major difference between birds and mammals lies in the expression of CSF1R. We developed monoclonal antibodies against CSF1R[33] and a transgenic chicken line expressing reporter genes from the CSF1R promoter region.[48] Combined analysis using these resources revealed the exceptionally high level of expression of CSF1R on antigen-capturing cells (follicle-associated epithelium [FAE]) in the bursa and lymphoid follicles (Balic A. et al., forthcoming) and in the respiratory tract.[49] These cells are unique to birds, which unlike mammals, lack lymph nodes. They are the functional equivalent of microfold (M) cells in mammals, the major site of luminal antigen sampling and pathogen invasion in intestinal mucosa.[50] In mammals, CSF1R is not expressed by M cells, although CSF1R-dependent macrophages in the lamina propria of the intestine control M cell differentiation.[51] In a second contrast with mammals, we found that CSF1R is highly expressed by antigen-presenting dendritic cells, which are a prevalent cell population in the avian liver in addition to their well-recognized prevalence in bursa and spleen.[52] So, we suggest two nonexclusive explanations. One is that a class of pathogen-associated virulence determinants acts to block binding of CSF1 or IL34 in order to compromise innate immunity or the function of FAE. Such a pathogenicity determinant exists in the form of the immunomodulatory BARF1 viral protein, which binds to human CSF1R.[43] A second nonexclusive explanation is that a pathogen or pathogen-associated molecule binds to CSF1R to enable receptor-mediated internalization. CSF1R is expressed on the cell surface and upon ligand binding promotes endocytosis of the ligand, either CSF1 or IL34.[3] Hence CSF1R could provide a portal for pathogen invasion.

The secondary question is how evolution in CSF1R can occur without compromising the innate immune system. CSF1 and CSF1R knockout mutations in mice and rats[4,5] are macrophage deficient and have severe developmental abnormalities. This is also the case in zebra fish.[12] We have recently confirmed based upon CRISPR-mediated knockout in the germ line that the chicken CSF1R is also absolutely required for posthatch development (Balic A. and DAH, forthcoming). Accordingly, loss of function mutations in CSF1R that abolish CSF1 binding in birds are unlikely to be tolerated as homozygotes and in the absence of a heterozygous phenotype there would be no selection. If CSF1R is the target of pathogen selection, only changes that preserve ligand binding would be tolerated. Such variation may well change the affinity of interaction with one or other ligand, but since binding is effectively irreversible (the ligand is internalized and degraded) this would not necessarily have any phenotypic impact. Lineage-specific evolution of CSF1 and IL34 subsequent to receptor divergence could occur by drift or positive selection to increase ligand binding affinity with consequent loss of binding to ancestral forms of the receptor. Based upon the phylogenetic tree, and the conservation of the majority of the zfCSF1 ligand variants in this region across distantly related species (kiwi, emu, emperor penguin, golden eagle), we suggest that zfCSF1 resembles the ancestral avian CSF1 sequence and the ability of this ancestral form of CSF1 to bind the CSF1R in chickens has been preserved across evolution. By the same argument, CSF1 from large animals including humans could be considered the ancestral form that retains binding to rodent CSF1R. We have not carried out the same functional analysis of IL34. Any pathogen that targets CSF1R would likely impact on both CSF1 and IL34 actions. Some contact amino acids in the receptor are shared within the hypervariable region of D2. The alignment in Supplementary Table 1B and analysis in Table 1 reveal that there are corresponding changes in Site 1 of IL34. Among the avian sequences we have analyzed, variants in Site 1 distinguish the galliforms (chicken, quail, and turkey) from all the other species.

Previous studies of birds in smallholder systems in Ethiopia provided strong evidence for heritable disease resistance and resilience.[45] Comparative analysis of available sequences of western commercial and tropically adapted populations identified prevalent protein sequence variants (Supplementary Tables 2 and 3). Some CSF1R variants distinguished layer and broiler lines consistent with evidence of signatures of selection in broiler lines in this region of chromosome 13[44] and QTL association with growth-related traits. CSF1R is clearly highly polymorphic in chickens and the coding variants distinguish western commercial birds from tropically adapted birds. By contrast, the much larger exome sequence database for humans (exac.broadinstitute.org) reveals 305 missense variants in CSF1R, but all except H362R (MAF = 0.19) are rare (MAF < 0.01). H362 is

not conserved in other mammals. Amino acid 362 lies within the receptor dimerization domain D4. Surprisingly, the H362R variant is strongly overrepresented (MAF 0.4) in East Asian populations and was associated with reduced receptor dimerization and altered CSF1 responsiveness.[53] Conversely, numerous synonymous coding variants and noncoding intronic variants in human CSF1R have MAF > 0.3. Two common variants that distinguish commercial broilers and layers, A308T and S409L, also occur within domain 4 but whether they influence CSF1R function is unknown. Common variants detected in commercial birds are relatively rare in Ethiopian and Nigerian populations and one CSF1R variant D91N was prevalent and unique to Ethiopia and Nigeria. Each of the variants affects an amino acid that is conserved to some degree across avian species. Polymorphism is a common feature of innate immune receptors.[54] It remains to be determined whether any of these variants can be associated with disease resistance or production traits and could represent targets for marker-assisted selection.

Although the focus of this study has been on the avian CSF1R system, as mentioned in the Introduction, there is emerging interest in CSF1R as a drug target[2] and in functional analysis of loss-of-function and gain-of-function mutations in *CSF1R* in human patients.[9–12] The human and mouse equivalents of the BaF3-CSF1R we have used here to assess cross-reaction of avian CSF1 have previously used to assay function of disease-associated human mutant receptors.[9] Our findings in birds suggest that focused mutagenesis of the interaction sites of CSF1 with CSF1R could provide the basis for generation of monomeric antagonists or higher affinity agonists.

AUTHORSHIP

D.A.H. performed informatic analysis, wrote the manuscript and supervised the study. M.W.G. performed structure and evolutionary analysis, M.W.G. and A.G. performed experiments, A.A.G. performed DNA sequence analysis. A.K., O.B., A.V.T. and J.S. collected samples, generated DNA sequence and contributed to analysis of data.

ACKNOWLEDGMENTS

The authors thank Dr. Andrew C. Gill for his help in producing recombinant domain swaps. D.A.H. is supported by the Mater Foundation. Whole genome sequencing was carried out by Edinburgh Genomics (Roslin, UK). This sequencing was funded in part by the Bill and Melinda Gates Foundation and the UK Government's Department for International Development (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the University of Edinburgh, SRUC (Scotland's Rural College), and the International Livestock Research Institute. The authors appreciate leadership and advice from Olivier Hanotte and Tadelle Dessie. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation nor the UK Government.

DISCLOSURES

The authors declare no conflicts of interest.

ORCID

David A. Hume https://orcid.org/0000-0002-2615-1478

REFERENCES

1. Chitu V, Stanley ER. Colony-stimulating factor-1 in immunity and inflammation. *Curr Opin Immunol*. 2006;18:39-48.
2. Hume DA, MacDonald KP. Therapeutic applications of macrophage colony-stimulating factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling. *Blood*. 2012;119:1810-1820.
3. Stanley ER, Chitu V. CSF-1 receptor signaling in myeloid cells. *Cold Spring Harb Perspect Biol*. 2014;6. 10.1101/cshperspect.a021857
4. Dai XM, Ryan GR, Hapel AJ, et al. Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood*. 2002;99:111-120.
5. Pridans C, Raper A, Davis GM, et al. Pleiotropic impacts of macrophage and microglial deficiency on development in rats with targeted mutation of the Csf1r locus. *J Immunol*. 2018;201:2683-2699.
6. Gow DJ, Sauter KA, Pridans C, et al. Characterisation of a novel Fc conjugate of macrophage colony-stimulating factor. *Mol Ther*. 2014;22:1580-1592.
7. Pridans C, Sauter KA, Irvine KM, et al. Macrophage colony-stimulating factor increases hepatic macrophage content, liver growth, and lipid accumulation in neonatal rats. *Am J Physiol Gastrointest Liver Physiol*. 2018;314:G388-G398.
8. Sauter KA, Waddell LA, Lisowski ZM, et al. Macrophage colony-stimulating factor (CSF1) controls monocyte production and maturation and the steady-state size of the liver in pigs. *Am J Physiol Gastrointest Liver Physiol*. 2016;311:G533-G547.
9. Pridans C, Sauter KA, Baer K, Kissel H, Hume DA. CSF1R mutations in hereditary diffuse leukoencephalopathy with spheroids are loss of function. *Sci Rep*. 2013;3:3013.
10. Rademakers R, Baker M, Nicholson AM, et al. Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids. *Nat Genet*. 2011;44:200-205.
11. Guo L, Bertola DR, Takanohashi A, et al. Bi-allelic CSF1R mutations cause skeletal dysplasia of dysosteosclerosis-pyle disease spectrum and degenerative encephalopathy with brain malformation. *Am J Hum Genet*. 2019;104:925-935.
12. Oosterhof N, Chang IJ, Karimiani EG, et al. Homozygous mutations in CSF1R cause a pediatric-onset leukoencephalopathy and can result in congenital absence of microglia. *Am J Hum Genet*. 2019;104:936-947.
13. Albagha OM, Visconti MR, Alonso N, et al. Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. *Nat Genet*. 2010;42:520-524.
14. Lin H, Lee E, Hestir K, et al. Discovery of a cytokine and its receptor by functional screening of the extracellular proteome. *Science*. 2008;320:807-811.
15. Wang Y, Szretter KJ, Vermi W, et al. IL34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat Immunol*. 2012;13:753-760.
16. Wei S, Nandi S, Chitu V, et al. Functional overlap but differential expression of CSF-1 and IL34 in their CSF-1 receptor-mediated regulation of myeloid cells. *J Leukoc Biol*. 2010;88:495-505.
17. Garceau V, Smith J, Paton IR, et al. Pivotal Advance: avian colony-stimulating factor 1 (CSF-1), interleukin-34 (IL34), and CSF-1 receptor genes and gene products. *J Leukoc Biol*. 2010;87:753-764.

18. Wang T, Kono T, Monte MM, et al. Identification of IL34 in teleost fish: differential expression of rainbow trout IL34, MCSF1 and MCSF2, ligands of the MCSF receptor. *Mol Immunol*. 2013;53:398-409.
19. Hume DA, Wollscheid-Lengeling E, Rojo R, Pridans C. The evolution of the macrophage-specific enhancer (Fms intronic regulatory element) within the CSF1R locus of vertebrates. *Sci Rep*. 2017;7:17115.
20. Garceau V, Balic A, Garcia-Morales C, et al. The development and maintenance of the mononuclear phagocyte system of the chick is controlled by signals from the macrophage colony-stimulating factor receptor. *BMC Biol*. 2015;13:12.
21. Chen X, Liu H, Focia PJ, Shim AH, He X. Structure of macrophage colony stimulating factor bound to FMS: diverse signaling assemblies of class III receptor tyrosine kinases. *Proc Natl Acad Sci USA*. 2008;105:18267-18272.
22. Felix J, Elegheert J, Gutsche I, et al. Human IL34 and CSF-1 establish structurally similar extracellular assemblies with their common hematopoietic receptor. *Structure*. 2013;21:528-539.
23. Ma X, Lin WY, Chen Y, et al. Structural basis for the dual recognition of helical cytokines IL34 and CSF-1 by CSF-1R. *Structure*. 2012;20:676-687.
24. Ellegren H. Comparative genomics and the study of evolution by natural selection. *Mol Ecol*. 2008;17:4586-4596.
25. Hughes AL, Friedman R. Codon-based tests of positive selection, branch lengths, and the evolution of mammalian immune system genes. *Immunogenetics*. 2008;60:495-506.
26. Laurin-Lemay S, Philippe H, Rodrigue N. Multiple factors confounding phylogenetic detection of selection on codon usage. *Mol Biol Evol*. 2018;35:1463-1472.
27. Yang Z, Nielsen R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol Biol Evol*. 2008;25:568-579.
28. Zhang G, Li C, Li Q, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 2014;346:1311-1320.
29. Botero-Castro F, Figuet E, Tilak MK, Nabholz B, Galtier N. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. *Mol Biol Evol*. 2017;34:3123-3131.
30. Gutowska MW. 2015. Comparative evolutionary and structural analysis of the avian and mammalian CSF1R systems. PhD Thesis, University of Edinburgh. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.685777.
31. Gheyas AA, Boschiero C, Eory L, et al. Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Res*. 2015;22:205-217.
32. Kranis A, Gheyas AA, Boschiero C, et al. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*. 2013;14:59.
33. Garcia-Morales C, Rothwell L, Moffat L, et al. Production and characterisation of a monoclonal antibody that recognises the chicken CSF1 receptor and confirms that expression is restricted to macrophage-lineage cells. *Dev Comp Immunol*. 2014;42:278-285.
34. Horiuchi K, Miyamoto T, Takaishi H, et al. Cell surface colony-stimulating factor 1 can be cleaved by TNF-alpha converting enzyme or endocytosed in a clathrin-dependent manner. *J Immunol*. 2007;179:6715-6724.
35. Nandi S, Akhter MP, Seifert MF, Dai XM, Stanley ER. Developmental and functional significance of the CSF-1 proteoglycan chondroitin sulfate chain. *Blood*. 2006;107:786-795.
36. Parmar HB, Barry C, Kai F, Duncan R. Golgi complex-plasma membrane trafficking directed by an autonomous, tribasic Golgi export signal. *Mol Biol Cell*. 2014;25:866-878.
37. Dib K, Tikhonova IG, Ivetic A, Schu P. The cytoplasmic tail of L-selectin interacts with the adaptor-protein complex AP-1 subunit mu1A via a novel basic binding motif. *J Biol Chem*. 2017;292:6703-6714.
38. Zur Y, Rosenfeld L, Bakhman A, et al. Engineering a monomeric variant of macrophage colony-stimulating factor (M-CSF) that antagonizes the c-FMS receptor. *Biochem J*. 2017;474:2601-2617.
39. Gow DJ, Garceau V, Kapetanovic R, et al. Cloning and expression of porcine colony stimulating factor-1 (CSF-1) and colony stimulating factor-1 receptor (CSF-1R) and analysis of the species specificity of stimulation by CSF-1 and interleukin 34. *Cytokine*. 2012;60:793-805.
40. Walter M, Lucet IS, Patel O, et al. The 2.7 A crystal structure of the autoinhibited human c-Fms kinase domain. *J Mol Biol*. 2007;367:839-847.
41. Prum RO, Berv JS, Dornburg A, et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. 2015;526:569-573.
42. Liu H, Leo C, Chen X, et al. The mechanism of shared but distinct CSF-1R signaling by the non-homologous cytokines IL34 and CSF-1. *Biochim Biophys Acta*. 2012;1824:938-945.
43. Felix J, De Munck S, Verstraete K, et al. Structure and assembly mechanism of the signaling complex mediated by human CSF-1. *Structure*. 2015;23:1621-1631.
44. Stainton JJ, Haley CS, Charlesworth B, Kranis A, Watson K, Wiener P. Detecting signatures of selection in nine distinct lines of broiler chickens. *Anim Genet*. 2015;46:37-49.
45. Psifidi A, Banos G, Matika O, et al. Genome-wide association studies of immune, disease and production traits in indigenous chicken ecotypes. *Genet Sel Evol*. 2016;48:74.
46. Nandi S, Cioce M, Yeung YG, et al. Receptor-type protein-tyrosine phosphatase zeta is a functional receptor for interleukin-34. *J Biol Chem*. 2013;288:21972-21986.
47. Segaliny AI, Brion R, Mortier E, et al. Syndecan-1 regulates the biological activities of interleukin-34. *Biochim Biophys Acta*. 2015;1853:1010-1021.
48. Balic A, Garcia-Morales C, Vervelde L, et al. Visualisation of chicken macrophages using transgenic reporter genes: insights into the development of the avian macrophage lineage. *Development*. 2014;141:3255-3265.
49. Sutton K, Costa T, Alber A, et al. Visualisation and characterisation of mononuclear phagocytes in the chicken respiratory tract using CSF1R-transgenic chickens. *Vet Res*. 2018;49. 10.1186/s13567-018-0598-7
50. Jang MH, Kweon MN, Iwatani K, et al. Intestinal villous M cells: an antigen entry site in the mucosal epithelium. *Proc Natl Acad Sci USA*. 2004;101:6110-6115.
51. Sehgal A, Donaldson DS, Pridans C, Sauter KA, Hume DA, Mabbott NA. The role of CSF1R-dependent macrophages in control of the intestinal stem-cell niche. *Nat Commun*. 2018;9:1272.
52. Hu T, Wu Z, Bush SJ, et al. Characterization of subpopulations of chicken mononuclear phagocytes that express TIM4 and CSF1R. *J Immunol*. 2019;202:1186–1199.
53. Yeh YM, Hsu SJ, Lin PC, et al. The c.1085A>G genetic variant of CSF1R gene regulates tumor immunity by altering the proliferation, polarization, and function of macrophages. *Clin Cancer Res*. 2017;23:6021-6030.
54. Villasenor-Cardoso MI, Ortega E. Polymorphisms of innate immunity receptors in infection by parasites. *Parasite Immunol*. 2011;33:643-653.

SUPPORTING INFORMATION

Additional information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Hume DA, Gutowska-Ding MW, Garcia-Morales C, et al. Functional evolution of the colony-stimulating factor 1 receptor (CSF1R) and its ligands in birds. *J Leukoc Biol*. 2019;1-14. https://doi.org/10.1002/JLB.6MA0519-172R